Comparison of Generalized Born and Poisson Models: Energetics and Dynamics of HIV Protease

LAURENT DAVID, RAY LUO, MICHAEL K. GILSON

Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received 27 July 1999; accepted 19 October 1999

ABSTRACT: This study characterizes the accuracy of energies and forces computed with a generalized Born (GB) model and the distance-dependent dielectric (DDD) model with respect to detailed finite solutions of the Poisson equation (FDPE). Tests are done for a small molecule in solution and for HIV-1 protease with inhibitor, KNI-272. GB agrees well with FDPE for the small molecule, but less well for the protein system. The correlation between GB and FDPE energies is poorest in calculations of changes upon binding. Also, forces computed with the GB model are less accurate than energies. The DDD model is far less accurate than GB. Nanosecond stochastic dynamics simulations of HIV-1 protease with an empty active site are used to examine the consequence of the models for the conformational preferences of the active site. Interestingly, the active site flaps remain near their starting conformations in the FDPE and GB simulations but collapse into the active site in the DDD simulation. J Comput Chem 21: 295–309, 2000

Keywords: generalized Born; Poisson equation; stochastic dynamics; HIV protease; implicit solvent models

Correspondence to: M. K. Gilson; e-mail: gilson@carb.nist.gov

Contract/grant sponsor: National Institutes of Health; contract/grant number: GM54053

Contract/grant sponsor: National Institute of Standards and Technology

Introduction

Biomolecules typically exist in an aqueous environment that has profound effects on their properties and functions. Biomolecular simulations must therefore account for the influence of the solvent if any degree of realism is to be achieved. An elegant and detailed approach is to include a large number of water molecules explicitly in the simulation, but this approach is computationally demanding. A number of laboratories have therefore developed implicit water models that aim to capture the effect of water at less computational cost.[1–12] The treatment of electrostatic interactions in implicit water models has proven particularly challenging because of the multibody nature of electrostatic interactions in an inhomogeneous dielectric environment.[13]

The multibody nature of electrostatic interactions can be captured by numerical solutions of the classical Poisson or Poisson–Boltzmann equations,[13–19] and this approach has gained acceptance as a convenient alternative to explicit representations of the solvent in energy calculations. The method can also be used to compute the electrostatic forces on atoms,[20–22] and can thus be incorporated into molecular dynamics simulations.[20, 23–26] However, very few simulations using numerical solutions of the Poisson–Boltzmann equation have been described in the literature, so it is difficult to evaluate the utility of this approach.

More recently, generalized Born (GB) models[3, 27–35] have emerged as rapid approximations to numerical solutions of the Poisson equation. Our applications of a GB model[3] in which atomic self-energies are estimated with a charge-induced dipole term,[36] have yielded good agreement with experiment for pK_a shifts in small molecules,[37] ion-pairing in aqueous solution,[38] the binding affinities of small molecules in chloroform,[39] conformational analysis of a potent HIV protease inhibitor,[40] and the association of adenine with synthetic adenine receptors (article in preparation). On the other hand, the GB model was not particularly effective at selecting the correct bound conformation of protein–ligand complexes,[41] and it gave somewhat unsatisfying results in calculations of the pK_as of ionizable groups in a protein active site.[37] Thus, the GB models already appear to be useful for small molecules but may require further evaluation and development for macromolecules. It is of particular interest to consider their use in the calculation of protein–ligand binding affinities, given the implications for structure-based drug design.

The present article is a study of continuum electrostatics models. The first section quantifies the accuracy of energies and forces computed with a GB model, relative to detailed finite difference solutions of the Poisson equation, for a small molecule and for HIV-1 protease with the potent inhibitor, KNI-272.[42] The accuracies of a distance-dependent dielectric model and of coarse-grid finite-difference calculations are also assessed. The second part of the study examines the use of continuum electrostatics models in dynamics simulations. It describes and compares three 1-ns molecular dynamics simulations of the HIV-1 protease active site, which differ only in their treatment of electrostatics: one uses finite-difference solutions of the Poisson equation; the second uses the GB model; and the third uses the distance-dependent dielectric model.

Methods

This work compares three different treatments of electrostatic solvation: the finite-difference Poisson equation (FDPE); the generalized Born (GB); and the distance-dependent dielectric constant $\epsilon_{ij} = 4r_{ij}$ (DDD). Energies and forces computed with the various electrostatics models are compared for static conformations of a small molecule and of a protein–ligand system, and stochastic dynamics simulations are carried out with the three models for HIV-1 protease with an empty active site. The following subsections detail the methods used in these calculations.

ENERGY COMPARISONS

Molecular structures. Energy comparisons are reported for two different molecular systems. The first is an isolated isoleucine residue. The program QUANTA[43] was used to add all hydrogens appropriate to the zwitterionic form of the residue in the conformation of Ile 62 of HIV-1 PR in the complex of HIV-1 PR with inhibitor KNI-272 (1hpx[44]). The molecule was then energy minimized with 100 steps of steepest descents, and a 100-ps stochastic dynamics calculation (see later) at 300 K was used to generate 200 conformations with a 0.5-ps interval. GB electrostatics was used in both the minimization and the dynamics. The electrostatic solvation energies of these conformations were computed with several approaches, and the results compared as described below.

The second molecular system consists of the active site of HIV-1 PR complexed with KNI-272. Solvent molecules and hydrogen atoms were deleted, and QUANTA was used to select those residues having any atom within 7 Å of any atom of KNI-272. All hydrogens were then added back for KNI-272, whereas only polar hydrogens were included for the protein. Aspartic acid 25A was protonated and Asp 25B was deprotonated, based on the NMR analysis of Wang et al.[45] The resulting system has 574 atoms, 87 for KNI-272 and 487 for the protein fragment. With Cα atoms fixed and no solvation model, 100 steps of steepest descent energy minimization were executed. One hundred different conformations of the complex were then generated by 100 ps of stochastic dynamics under the same conditions, with a snapshot saved every 1 ps. Each snapshot $i \in [1, 100]$ contains coordinates for the active site residues of HIV-1 PR and for KNI-272. Various approaches were used to compute the electrostatic solvation energies of the entire complex ($\Delta G_{\text{complex},i}$), the protein without the inhibitor ($\Delta G_{\text{protein},i}$), and the inhibitor without the protein ($\Delta G_{\text{inhibitor},i}$). The change in electrostatic solvation energy upon binding, for each snapshot, also was computed as $\Delta G_{\text{bind},i} \equiv \Delta G_{\text{complex},i} - \Delta G_{\text{protein},i} - \Delta G_{\text{inhibitor},i}$.

Calculation of electrostatic energies. Electrostatic models were compared against reference calculations with the FDPE method with a fine grid spacing of 0.25 Å, which necessitated the use of a 200^3 grid for the protein–ligand complex. For both the isoleucine and the protein–ligand cases, three models were compared with the reference FDPE calculations: GB; FDPE with a coarse grid (1 Å spacing); and DDD. For the protein–ligand case, two additional models were examined: a reparameterized form of GB, and FDPE with a very coarse grid spacing (2 Å).

The base implementation of the GB model used was that of Qiu et al.[31] For the protein–ligand system, we also sought to improve the agreement between GB and the reference FDPE results by adjusting parameters P1–P5 from Qiu et al. Three new parameter sets were generated, one each for KNI-272, HIV-1 PR, and the complex HIV-1 PR/KNI-272. The training set for each optimization consisted of every 20th conformation of the 100 test conformations just described, for a total of five training conformations for each molecular species. The parameters were optimized with the same simulated annealing/simplex method employed in the original publication[31] and reimplemented locally.

In the DDD model, the total electrostatic energy is computed as a sum of pairwise interactions between atomic charges i and j with an effective dielectric constant $\epsilon_{ij} = 4r_{ij}$, where r_{ij} is the interatomic distance. In the absence of solvent, a dielectric constant of 1 would be used. We therefore compute the electrostatic solvation energy associated with the DDD model as:

$$\Delta G_{\text{DDD}} = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \left(\frac{1}{4r_{ij}} - 1 \right) \tag{1}$$

FORCE COMPARISONS

Molecular structures. A single molecule with N atoms yields $3N$ force components, which is enough to establish a meaningful comparison between different models. Therefore, only two molecular structures were used to assess the accuracy of the various electrostatic models against the reference FDPE method. One structure is that of Ile 62A extracted from the HIV-1 PR structure, with all hydrogens added and both termini charged. This case provides information on the electrostatic forces for highly solvated atoms. The other structure is the crystal conformation of the HIV-1 PR/KNI-272 complex with the KNI-272 removed. Forces are computed only for those atoms treated as mobile in the stochastic dynamics simulation of this complex (see Dynamics Comparisons subsection).

Calculation of electrostatic forces. Detailed finite difference solutions of the Poisson equation were used to obtain reference values for the electrostatic forces due to solvent. These FDPE forces may be separated into two parts.[20–22, 46] One results from the interaction of each atomic charge q_i with the reaction field E_i^{RF} at the atom generated by the solvent, and the other, F_i^{DBF}, results from the inward pressure of the solvent on the dielectric boundary. This is a consequence of the tendency of the high-dielectric solvent to reduce the electrostatic energy by displacing low-dielectric solute. Thus:

$$F^{\text{elec,solv}} = q_i E_i^{\text{RF}} + F_i^{\text{DBF}} \tag{2}$$

The method used to compute these has been described previously.[22, 23] Calculating the dielectric boundary force requires calculating the derivative of the dielectric constant at each grid line with respect to the positions of nearby atoms.[22] These derivatives become nonphysically large in the rare event that an atom is tangent to the center of a grid branch [see eq. (22) in ref. 22]. This numerical problem is avoided here by setting the dielectric boundary force associated with such grid branches

to zero when the distance in the denominator of eq. (22) of ref. 22 is less than 0.05 Å.

The FDPE computation is done in two stages. The first stage uses a large, coarse grid that accomodates the entire protein. The second stage uses a small, finely spaced grid that encompasses the atoms whose forces are to be calculated, as well as an additional boundary zone approximately 7 Å thick. The potential at the boundary of the fine grid is interpolated from the coarse grid; the coarse grid is only used to define the boundary conditions of the fine grid. The reference calculations use a $30 \times 30 \times 30$ coarse grid with spacing 3.5 Å and a $280 \times 180 \times 204$ fine grid with spacing 0.25 Å. The finite-difference calculations are iterated until the fractional change in the norm of the residual falls below 10^{-6}. The sensitivity of the results to the spacing of the fine grid were examined by additional FDPE calculations that use a grid spacing of 1.0 Å.

Two sets of GB force calculations are compared with the reference FDPE forces. In the first set, the total derivative of the electrostatic energy with respect to atomic coordinates is computed; these calculations include the differentials of the effective Born radii with respect to atomic coordinates. Computing these differentials makes the method rather time-consuming: the computational work varies as the cube of the number of atoms. Therefore, we also examine the time-saving simplification of setting the differentials of the effective Born radii to zero. With this approximation, the time required to calculate the GB forces varies as the square of the number of atoms. The GB calculations of atomic forces use the published parameters P1–P5 of Qiu et al.[31]

The distance-dependent dielectric approximation, $\epsilon_{ij} = 4r_{ij}$, is also examined. Here, ϵ_{ij} is the effective dielectric constant for the interaction of atoms i and j, and r_{ij} is the interatomic distance. The full differential of the energy is used; thus, the distance dependence of ϵ is included in the derivative. The Coulombic part of the electrostatic force is subtracted from the total force computed with the distance-dependent model to leave only the solvent contribution, which is directly comparable with the FDPE results.

DYNAMICS COMPARISONS

Three stochastic dynamics simulations were carried out: one obtains solvation forces from finite-difference solutions of the Poisson equation; one uses the generalized Born approximation; and one uses a distance-dependent dielectric model. Otherwise, the calculations are the same.

Molecular structures. All three simulations use the same initial structure of HIV-1 protease with an empty active site. The initial conformation was drawn from the crystal structure of HIV-1 PR with the inhibitor KNI-272,[40] but the inhibitor was deleted. Polar hydrogen coordinates were reconstructed with the program QUANTA.[43] All carboxylic acids, lysines, and arginines were treated as ionized for a net enzyme charge of +4 elementary charges. (It is appropriate to treat the active site aspartyl groups as ionized for the free enzyme at pH 7.0.[47]) The structure was relaxed by 100 steps of steepest descents energy minimization with the appropriate electrostatics model, with only the atoms treated as mobile in the simulations allowed to move.

Stochastic dynamics method. Stochastic dynamics simulations[48] of 1-ns duration were used to compare the dynamical and structural consequences of the FDPE, GB, and DDD treatments of electrostatic interactions. The simulations focus on the dynamics of the active site "flaps" of HIV-1 PR with no ligand. The flaps are believed to adopt an open conformation in the absence of ligand,[49] so we expected that they would be unstable in their starting conformation and that the choice of electrostatic model might significantly influence their motions. To speed the calculations, only the flaps and their hinge regions were treated as mobile; this corresponds to residues 36 to 60 for the elbows and tips of the flaps, and residues 77 to 83 for the loops covering the side of the flap. The same residues were treated as mobile in both subunits, for a total of 600 mobile atoms (see Fig. 1).

The temperature was set to 300 K in all three simulations and the friction coefficient in the stochastic dynamics algorithm is set to 10 ps^{-1}. The integration timestep was 1.0 fs and a coordinate set was stored every 200 fs, for a total of 5000 frames per 1-ns simulation. Temperatures were computed with the standard equation:

$$T = \frac{2\langle KE \rangle}{k_B 3 N_{\text{mobile}}} \tag{3}$$

where $\langle KE \rangle$ is the time-averaged kinetic energy, k_B is the Boltzmann constant, and N_{mobile} is the number of mobile atoms, which is one third of the number of degrees of freedom.

In the GB and DDD simulations, an atom-based cutoff of 13 Å was used for all Lennard–Jones and electrostatic interactions, and the pairlist was updated every 25 fs. The same pairlist was used in computing the effective atomic radii and the electrostatic solvation energy for the GB simulations. The

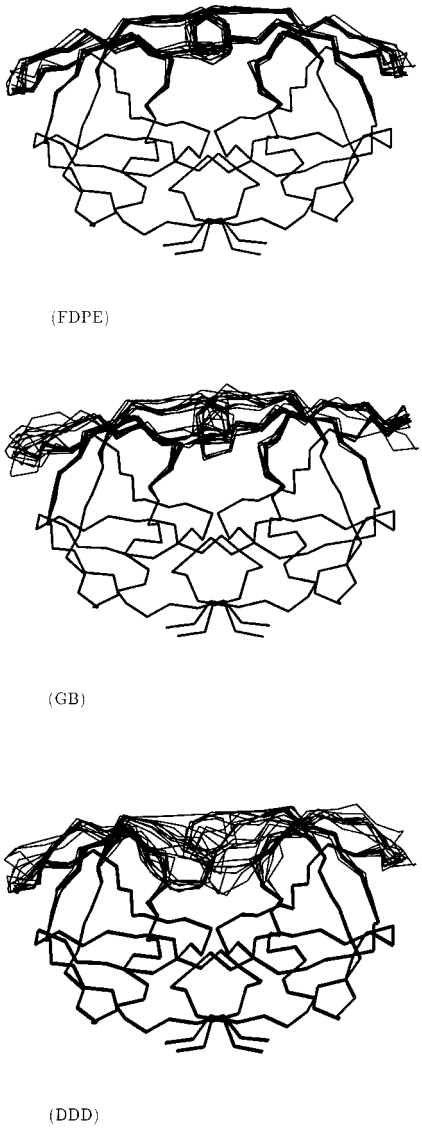


FIGURE 1. Superposition of 10 snapshots (every 100 ps) for the FDPE (top), GB (middle), and DDD (bottom) simulations.

treatment of nonbonded interactions in the FDPE simulations is as follows.

Calculation of electrostatic forces. The GB implementation in the dynamics calculations uses the published parameters (P1–P5)[31] and the time-saving simplification of neglecting the differentials of the effective Born radii with respect to atomic coordinates (see above). The calculation of forces with the distance-dependent dielectric model was as described in the Dynamics Comparisons subsection.

The FDPE solvation forces were also computed essentially as described in the Dynamics Comparisons subsection. However, fewer iterations were executed in the finite-difference calculations in order to save computer time. As shown in Table I, the forces computed with a convergence criterion of 10^{-2} agreed extremely well with those computed with the conservative criterion of 10^{-6}, and they required fivefold fewer iterations. The value of 10^{-2} was therefore used in the present simulations. During the simulation, the coarse and fine grid potentials were updated every 6 fs and 3 fs, respectively.

The electrostatic fields obtained with the FDPE method account not only for interactions of charged solute atoms with the solvent, but also the Coulombic interactions among atoms of the solute. Thus, the field at atom i may be written as $E_i = E_i^{RF} + E_i^{coul}$. However, it is not appropriate to use the grid to compute short-range Coulombic interactions, for two reasons. First, the Coulombic interaction of two atoms that are close relative to the spacing of the FDPE grid is highly inaccurate. Second, the CHARMM22 force field dictates that Coulombic interactions between atoms in 1-2 and 1-3 bonding relationships be neglected entirely, and that Coulombic interactions between atoms in a 1-4 bonding relationship be scaled by a factor of 0.5.[50] Thus, only the long-ranged part of the Coulombic field obtained from the grid is useful in these simulations.

We address these problems as follows. First, the fine grid is made large enough to encompass not only the mobile atoms, but also an additional buffer region extending 8 Å in each direction beyond a rectangular box enclosing the mobile atoms. Second, the Green's function method[51] is used to subtract from the FDPE solutions all potentials generated by atoms within the fine grid (see earlier). This removes all short-ranged Coulombic interactions involving the mobile atoms, leaving only the reac-

TABLE I.
Accuracy and Number of Iterations for FDPE Calculations as a Function of Convergence Tolerance (Conv), Analyzed via Linear Regressions Relative to Results for Conv = 10^{-6}.

log(Conv)	CC	Slope	*y*-Intercept	Iterations
−0.5	0.57	1.38	−0.009	5
−1	0.92	1.01	0.01	14
−1.5	0.998	1.0002	−0.0016	30
−2	0.99994	1.0006	0.0010	58
−3	0.9999998	0.999994	−0.000012	109
−4	1.0	1.000002	$2 \cdot 10^{-6}$	155
−6	1	1	0	254

CC, correlation coefficient.

tion field and the Coulombic field due to immobile atoms outside the fine grid. (The influence of these atoms is mediated by their influence on the boundary potential of the fine grid.) Finally, the inappropriate Coulombic interactions are substituted by correct Coulombic interactions obtained via a customary nonbonded pairlist with a 13-Å cutoff that accounts for the interactions among mobile atoms and between mobile atoms and those fixed atoms encompassed by the fine grid and further than 1 Å from the edge of the fine grid. Lennard–Jones interactions are computed with the same pairlist.

The lattice dimensions of the coarse grid are 30 × 30 × 30, with a spacing of 3.5 Å; this encompasses the entire protein. The lattice dimensions of the fine grid are 72 × 45 × 51, with a spacing of 1.0 Å. This grid extends over all mobile atoms and adds an 8-Å margin on each side.

Analysis of the trajectories. Summary information about the trajectories is obtained by computing root-mean-square deviations (RMSD) of protein conformations relative to the starting conformation, positional fluctuations of atoms with respect to their average positions for comparison with crystallographic Debye–Waller factors, positional cross-correlations among atoms, and the probabilities of specific hydrogen bonds. These quantities are obtained as follows.

The RMSD relative to the initial conformation is computed as:

$$\mathrm{RMSD} = \left(\frac{1}{N_{\mathrm{mobile}}} \sum_{i}^{N_{\mathrm{mobile}}} \left(x_i - x_i^0\right)^2 + \left(y_i - y_i^0\right)^2 + \left(z_i - z_i^0\right)^2 \right)^{1/2} \quad (4)$$

where N_{mobile} is the number of mobile atoms; x_i, y_i, z_i are the Cartesian coordinates of atom i; and 0 indicates coordinates of the initial conformation. It is not necessary to fit the new coordinates to the initial ones because the protein is not free to rotate or translate. No correction is made for the existence of multiple equivalent conformations of symmetrical, rotatable groups, such as phenylalanines or carboxylates.[41] However, our experience is that such corrections are well below 1 Å for molecules that are not symmetric under external rotation.

The positional fluctuation of atom i is computed as:

$$f_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{r}_i^2 \rangle \quad (5)$$

where $\Delta \mathbf{r}_i$ is the deviation of atom i from its mean position during the simulation and angle brackets indicate averages over the stored conformations. The factor of $8\pi^2$ makes the results comparable with crystallographic Debye–Waller factors.

The positional cross-correlation of atoms i and j is calculated as:[52, 53]

$$C_{ij} = \langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle / \left(\langle \Delta \mathbf{r}_i^2 \rangle \langle \Delta \mathbf{r}_j^2 \rangle \right)^{1/2} \quad (6)$$

In the analysis of hydrogen bonds, the criterion for existence of a hydrogen bond is a donor–acceptor distance of ≤3.2 Å. The probability of a hydrogen bond is computed as the fraction of stored conformations in which the bond exists, over a specified time interval.

PARAMETERS AND COMPUTATIONAL DETAILS

All calculations use CHARMM22 parameters,[50] with solute and solvent dielectric constants of 1 and 80, respectively. The ionic strength of the solvent is set to zero so that the results can be compared with the GB model, which does not account for a dissolved electrolyte. The dielectric boundary between solute and solvent in the FDPE calculations is defined as the Richards molecular surface[54] computed with a 1.4-Å probe sphere. This definition is consistent with that used in parameterizing the GB model.[31] The atomic cavity radii of hydrogens are set to 1.2 Å and the radii of the other elements are set to their Lennard–Jones R_{min} parameter.[50] Dielectric boundary smoothing is used in the FDPE calculations to increase the precision of the placement of the surface.[21] Except as otherwise noted, all calculations use a local version of the program UHBD.[18]

Results

COMPARISON OF ELECTROSTATIC SOLVATION ENERGIES

Tables II and III describe the accuracy of energies computed with the various electrostatic solvation models, based on comparisons with reference results from detailed finite-difference solutions of the Poisson equation. The structures considered in Table II come from a 100-ps stochastic dynamics simulation on a single isoleucine molecule. The structures considered in Table III come from a brief simulation of HIV-1 protease with KNI-272 *in vacuo*; results are presented for the complex, the separated molecules, and for the change upon "binding."

For isoleucine (Table II), GB agrees well with the reference results, and FDPE calculations with a coarse (∼1 Å) grid are less accurate than GB. The

TABLE II.
Accuracy of Electrostatic Solvation Energies (kJ/mol) for Isolated Isoleucine, Analyzed via Linear Regressions and Root-Mean-Squared Deviations Relative to Results with a (0.25-Å, 100^3) Grid.

	Slope	CC	*y*-Intercept	RMSD
GB	0.99	0.96	−1.65	1.40
FDPEC	1.09	0.94	1.97	2.81
DDD	1.18	0.89	148.1	138.9

CC, correlation coefficient; GB, generalized Born; DDD, distance-dependent dielectric; FDPEC, finite-difference Poisson equation with a coarse (1-Å, 25^3) grid; RMSD, root-mean-square deviation.

distance dielectric model is not very accurate, primarily because of the large *y*-intercept; the slope and correlation coefficients are surprisingly close to 1. The relatively poor accuracy obtained with a coarse FDPE grid is related to the fact that the system is highly solvated, and the coarse grid gives a poor description of the dielectric boundary.

For the HIV-1 protease system (Table III), the GB results correlate well with FDPE results for the complex and the separated molecules: the correlation coefficients are 0.91 to 0.96 and the slopes are close to 1. However, the linear regressions yield *y*-intercepts that deviate widely from zero by up to 190.1 kJ/mol. Also, the correlation becomes worse for the "binding energy": the correlation coefficient and slope become 0.87 and 0.82, respectively. Given the large intercepts, it is not surprising that the RMSDs are high (>20 kJ/mol).

The inaccuracy of the GB calculations motivated a reparameterization of the GB model for the molecules at hand. Three new sets of parameters (P1–P5) were derived, one for the complex, one for the protease alone, and one for KNI-272 alone. Five conformations of each species were used for the reparameterizations and the energies from the detailed FDPE calculations were used as target values. The three new sets of parameters that resulted are

TABLE III.
Accuracy of Electrostatic Solvation Energies (kJ/mol) for HIV-1 PR and KNI-272.

		Slope	CC	*y*-Intercept	RMSD
GB	$\Delta G_{complex}$	0.93	0.92	−313.1	190.0
	ΔG_{HIVP}	1.02	0.96	−55.8	94.3
	ΔG_{KNI}	1.03	0.91	27.8	22.3
	$\Delta\Delta G_{bind}$	0.87	0.82	−73.4	119.0
GB^{RP}	$\Delta G_{complex}$	0.96	0.95	−74.9	19.1
	ΔG_{HIVP}	1.07	0.97	120.0	19.5
	ΔG_{KNI}	1.01	0.90	4.28	2.9
	$\Delta\Delta G_{bind}$	1.02	0.84	−7.45	14.4
DDD	$\Delta G_{complex}$	0.84	0.84	7490.0	7760.0
	ΔG_{HIVP}	0.916	0.90	7400.0	7560.0
	ΔG_{KNI}	1.28	0.85	437.0	390.0
	$\Delta\Delta G_{bind}$	0.86	0.76	−142.0	194.0
FDPEC (1 Å)	$\Delta G_{complex}$	1.02	0.99	−62.3	104.0
	ΔG_{HIVP}	1.01	0.99	−102.0	115.0
	ΔG_{KNI}	1.06	0.95	2.64	6.87
	$\Delta\Delta G_{bind}$	0.98	0.87	25.4	20.7
FDPEC (2 Å)	$\Delta G_{complex}$	1.01	0.81	−188.0	204.0
	ΔG_{HIVP}	0.759	0.70	−684.0	235.0
	ΔG_{KNI}	1.28	0.53	18.3	30.8
	$\Delta\Delta G_{bind}$	1.48	0.421	−115.0	87.7

$\Delta G_{complex}$, HIV-1 PR complexed with KNI-272; ΔG_{HIVP}, HIV-1 PR alone; ΔG_{KNI}, KNI-272 alone; $\Delta\Delta G_{bind}$, change upon binding. GB, generalized Born; GB^{RP} reparameterized generalized Born; DDD, distance-dependent dielectric; FDPEC (1 Å), finite difference Poisson equation with 1-Å grid; FDPEC (2 Å), same with 2-Å grid. For reference, the ranges of solvation energies (kJ/mol) obtained with FDPE are: $\Delta G_{complex} \in [1536, 1824]$; $\Delta G_{HIVP} \in [1678, 2026]$; $\Delta G_{KNI} \in [158, 179]$; $\Delta\Delta G_{bind} \in [269, 403]$.

TABLE IV.
GB Parameters Reoptimized for HIV-1 PR Bound to KNI-272, HIV-1 PR Alone, and KNI-272 Alone.

Parameter	Complex	HIV-1 PR	KNI-272	Published
P1	0.122	0.0717	0.1096	0.073
P2	0.010	0.14	0.050	0.921
P3	0.526	0.527	0.7104	6.2105
P4	25.0	24.88	20.725	15.236
P5	0.864	1.03	1.102	1.254

Published original values from ref. 31.

given in Table IV. As shown in Table III (GB^{RP}), reparameterization markedly improved the accuracy of the GB model for all three systems and for the "binding energy."

The distance-dependent dielectric (DDD) yields results that are, for the most part, markedly less accurate than those from GB or GB^{RP}. Surprisingly, however, the net "binding energies" show correlation coefficients only slightly worse than those for GB: 0.76 vs. 0.82.

Finite-difference solutions of the Poisson equation with coarse grids—spacings of 1 Å and 2 Å—also are compared with the detailed reference calculations. The 1-Å grid spacing gives more accurate results than standard GB, and the slopes and correlation coefficients are better than those for reparameterized GB as well. However, the *y*-intercepts, and thus the RMSDs, are not as good as those for reparameterized GB.

COMPARISONS OF FORCES FOR STATIC STRUCTURES

Tables V and VI assess the accuracy of several methods of estimating solvent electrostatic forces, where the reference values are obtained by detailed finite-difference solutions of the Poisson equation. Table V uses a single conformation of isoleucine and Table VI uses a single conformation of the HIV-1 PR/KNI-272 complex. The statistics are based on $3N$ force components, where N is the number of atoms included for each structure. GB forces are computed with full gradients (GB^{full}), thus including the differentials of the atomic cavity radii with respect to atomic coordinates, and also with a simplified form (GB^{smpl}) that ignores these differentials and thus saves computer time. Forces also are computed with finite-difference solutions of the Poisson equation using a coarser grid spacing (1.0 Å) than for those used in the reference calculations (0.25 Å).

TABLE V.
Accuracy of Forces (kJ/mol·Å) Obtained from Electrostatic Solvation Models Relative to Specified Reference Models for Isoleucine Alone (66 Force Components).

Test Model	Reference Model	Slope	CC	RMSD
GB^{full}	$FDPE^{fine}$	0.92	0.94	5.3
GB^{smpl}	$FDPE^{fine}$	0.80	0.92	6.1
$FDPE^{coarse}$	$FDPE^{fine}$	0.96	0.98	3.1
DDD	$FDPE^{fine}$	0.78	0.83	9.1
GB^{smpl}	GB^{full}	0.88	0.98	3.1

GB^{full}, generalized Born with full derivatives; GB^{smpl}, generalized Born neglecting differentials of effective Born radii; $FDPE^{fine}$, FDPE with fine grid; $FDPE^{coarse}$, FDPE with coarse (1-Å) grid. The *y*-intercepts are essentially 0 in all cases.

For isoleucine, both the full and simplified GB models, GB^{full} and GB^{smpl}, yield good agreement with the reference finite difference calculations (Table V). The full gradient version of the GB forces is somewhat more accurate than the approximate version in which the differentials of atomic radii are neglected. (For completeness, Table V also contains statistics on the differences between the full GB and the simplified GB results.) The distance-dependent dielectric model is much less accurate than either GB implementation. However, results obtained with the coarse-grid FDPE calculation with a grid spacing of 1 Å are considerably more accurate than both GB models. This is surprising, given that the opposite result was obtained in the energy calculations described earlier. Perhaps this can be explained by the fact that the GB model was created and parameterized based on energies, not forces.

A similar pattern of results is observed for the HIV-1 PR/KNI-272 complex, but the GB models are less accurate than they are for isoleucine: the linear regression slopes deviate further from unity and the correlation coefficients fall to about 0.8 from about 0.9 in the isoleucine test case. Neglecting the differentials in computing the GB forces (GB^{smpl}) does not markedly degrade accuracy from the full computation of GB forces (GB^{full}). However, the simplification reduces the computer time by a factor of 400, making this an appealing approximation for use in molecular simulations. The performance of the distance-dependent dielectric model also is worse for HIV-1 PR/KNI-272, as is most evident in the correlation coefficient of 0.4.

Interestingly, when FDPE is used with a coarse grid spacing of 1.0 Å, the resulting forces still agree

TABLE VI.
Accuracy of Forces—Data Same as Table V but for HIV-1 PR with an Empty Active Site.

Test Model	Reference Model	Slope	CC	*y*-Intercept	RMSD
GB^{full}	FDPE (0.25)	0.78	0.81	−0.04	6.1
GB^{smpl}	FDPE (0.25)	0.74	0.78	−0.01	6.9
DDD	FDPE (0.25)	0.68	0.40	−0.08	42.0
FDPEC (1.0)	FDPE (0.25)	0.93	0.93	0.08	2.5
GB^{smpl}	GB^{full}	0.97	0.98	0.04	0.51

very well with the reference calculations with a fine grid of 0.25-Å spacing. However, the computer time required for FDPE with a 1.0-Å grid is only three times greater than that for simplified GB. Further optimization of both the finite-difference code and of the GB implementation might alter this balance, but this observation does suggest that further development of finite-difference methods for computing electrostatic solvation forces may be worthwhile, despite the simplicity of the GB approach.

It is also worth mentioning some test calculations not listed in the tables. These examine the influence of the initial, coarse focusing grid on the computed forces. First, when the initial focusing grid is neglected and the boundary potentials of the second (fine) grid are simply set to zero, the root-mean-square (RMS) change in the electrostatic solvation forces is 11 kJ/mol·Å. This large error implies that it is worth using the initial focusing calculation to establish realistic boundary potentials for the second, finer focusing grid. On the other hand, using an initial grid that is three-fold more closely spaced yields an RMS change in atomic forces of only 0.2 kJ/mol·Å, so a detailed grid is not required.

SIMULATIONS OF HIV-1 PR WITH AN EMPTY ACTIVE SITE

Three 1-ns simulations permit examination of the differences in the motions and conformational preferences that result from the various electrostatics models. One simulation uses the distance-dependent dielectric model. The second uses the simplified GB model that neglects the differentials of the effective Born radii. The third uses finite-difference solutions of the PB equation with a 1-Å grid spacing. These simulations required 4, 10, and 45 days, respectively, on a single-processor R10000 SGI Octane machine. The test system is HIV-1 PR with only the active site flaps treated as mobile. To elicit dynamics that might distinguish among the three models, the simulations were started with the crystal conformation of the HIV-1 PR/KNI-272 complex,[44] but with KNI-272 artificially deleted. Thus, the active-site flaps are started in a closed conformation that is expected to be unstable, because crystal structures of HIV-1 PR with an empty active site[49] show the flaps in an open conformation.

Structural analysis of the simulations. Figure 1 shows the loop conformations from each simulation at 100-ps intervals. Perhaps the most striking result is that both flaps collapse into the active site in the DDD simulation. This appears to be an incorrect result, given than the crystal structure of HIV-1 PR with no ligand shows the loops moving away from the active site and not deeper into the active site. The time course of the collapse can be seen in Figure 2, which plots the distance of the tips of the loops, defined by the α carbons of Ile 50A and Ile 50B, from the base of the active site, defined by the α carbons of Asp 25A and Asp 25B. Loop B collapses into the active site within 100 ps, whereas loop A collapses between 700 and 800 ps.

In contrast, the loops do not collapse in the FDPE and GB simulations. This is evident from the structures in Figure 1 and the graphs in Figure 2. Loop A does move gradually into the active site in the GB simulation over the 100- to 400-ps time interval, but it then drifts up again over the rest of the simulation.

The loop motions are associated with adjustments in hydrogen bonds linking loops A and B. For the GB and FDPE simulations, the following relatively stable hydrogen bonds are observed: Ile 50A NH—Gly 51B CO; Gly 51A NH—Gly 52A CO; and Ile 50A NH—Ile 50B CO. For the DDD simulation the most occupied hydrogen bonds are as follows: Gly 51A NH—Gly 49B CO; Gly 52A NH—Gly 49B CO; and Ile50A NH—Ile 50B CO. However, these hydrogen bonds are unstable in the DDD simulation; indeed, they disappear around 750 ps after the second flap collapses into the active site. For the GB and FDPE simulations, the hydrogen bonds between Gly 52A and Gly 51B and between Ile 50A and Ile 50B are highly occupied during the last

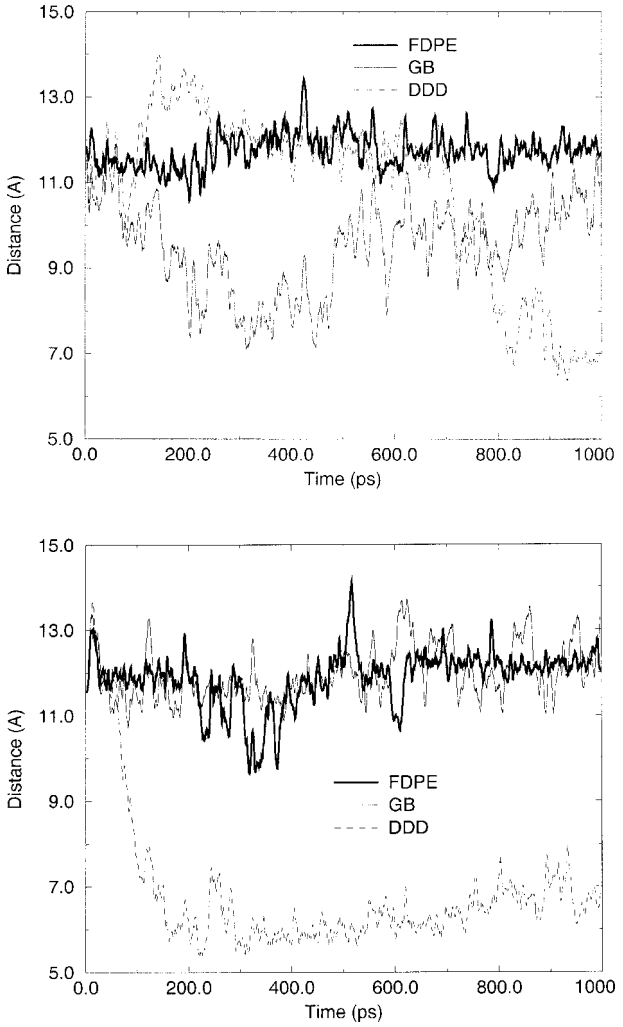


FIGURE 2. Distances between the tips of the flaps and the catalytic aspartyl groups. Top: monomer A (distance: Asp 25A–Ile 50A); bottom: monomer B (distance: Asp 25B–Ile 50B). Thick line: FDPE; thin line: GB; dashed line: DDD.

500 ps of both simulations. The only hydrogen bond between the two flaps in the crystal structure disappears at around 350 ps for the FDPE simulation and at around 900 ps for the GB simulation, and is essentially absent in the DDD simulation.

In summary, the FDPE and GB models yield rather similar conformational distributions for the flaps. The flaps do not adopt the open conformation seen crystallographically,[49] but the GB simulation in particular shows considerable mobility. In the DDD simulation, however, the flaps collapse into the active site, a result that contrasts sharply with the open conformation found in HIV-1 PR with no ligand[49] and with the other two simulations.

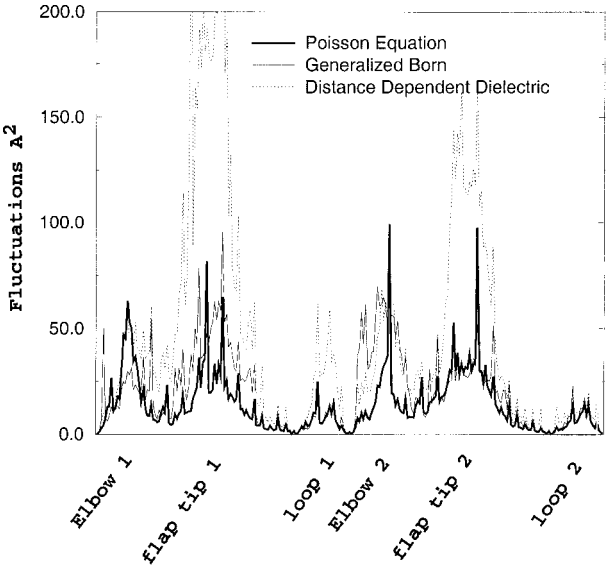


FIGURE 3. Positional fluctuations of nonhydrogen main-chain atoms in the three simulations.

Positional fluctuations. Figure 3 presents the positional fluctuations of the mobile main-chain atoms for all three simulations, computed relative to the structures obtained by averaging coordinates over the last 500 ps. (Those atoms with very small fluctuations are directly bonded to fixed atoms.) The tips of the active-site flaps are the most mobile parts of the protein, and the elbows are also relatively mobile. The DDD model leads to significantly greater mobility than the GB model, which in turn shows somewhat greater mobility than the FDPE

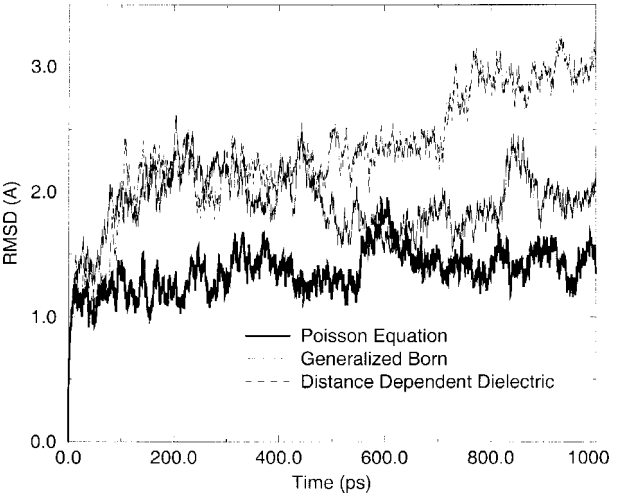
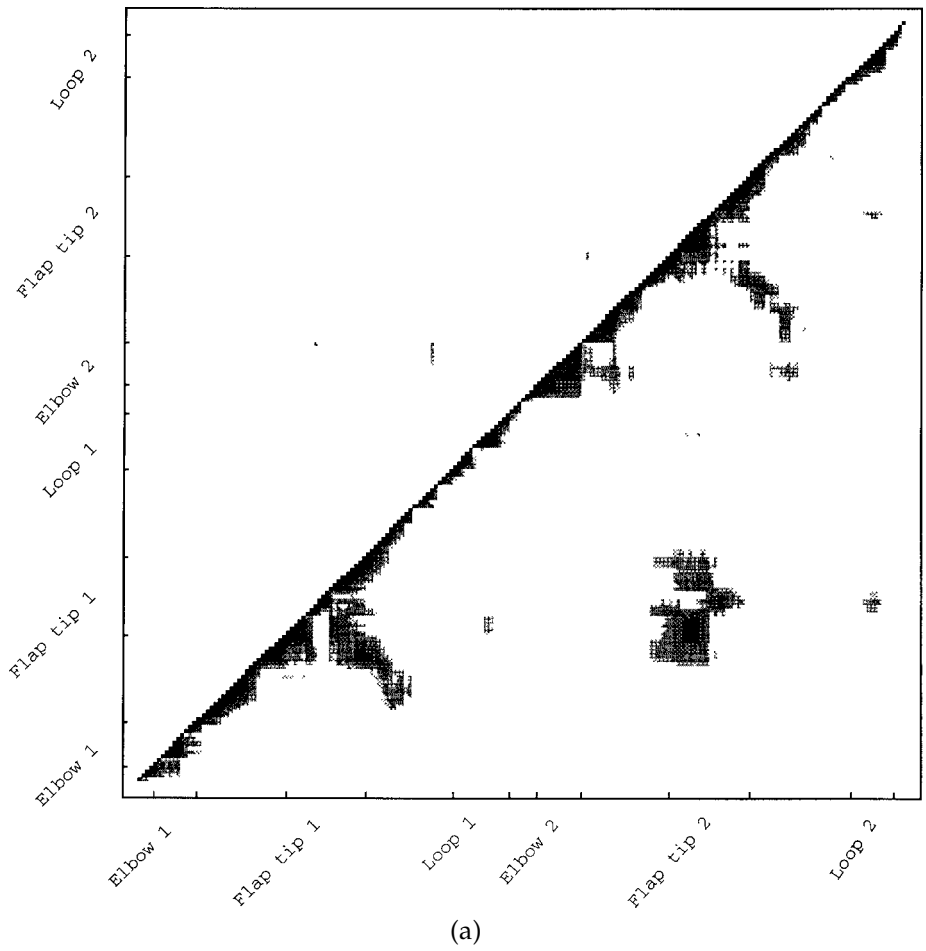


FIGURE 4. Root-mean-squared deviations of nonhydrogen main-chain atoms from the starting coordinates in the three simulations.



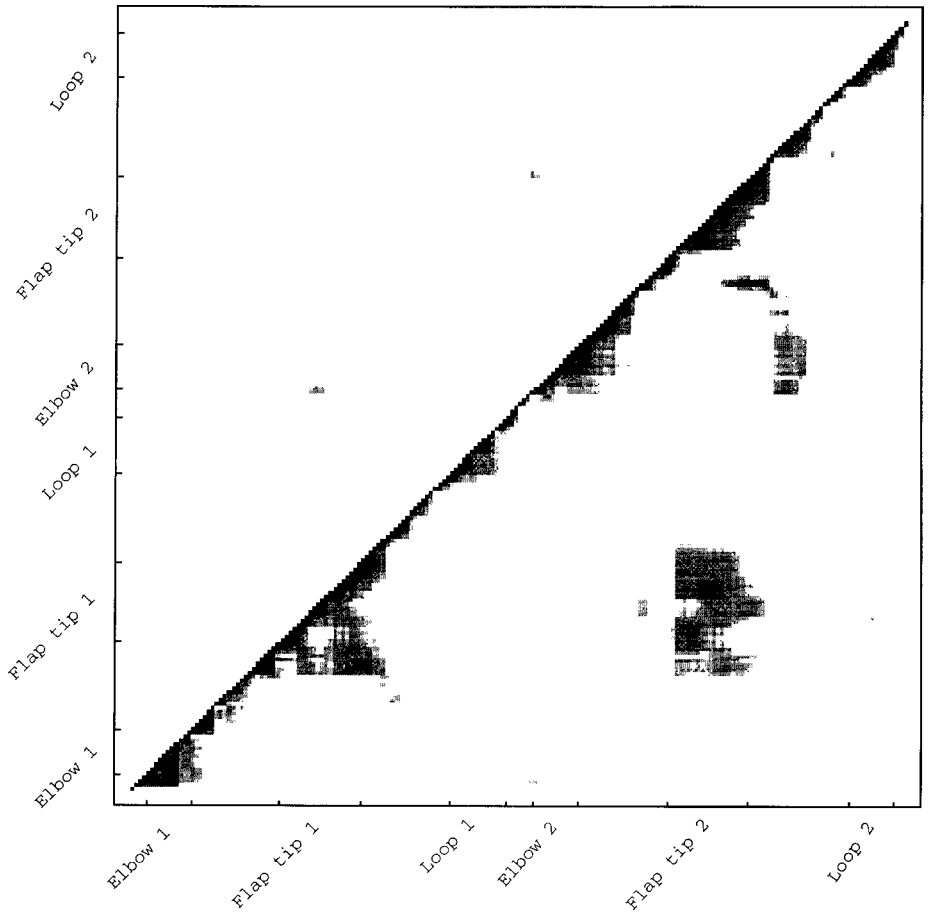
(a)

FIGURE 5. Cross-correlations of the positions of mobile nonhydrogen main-chain atoms in the three simulations. Positive and negative correlations are shown in the lower right and upper left triangles, respectively. A darker square indicates a stronger (anti)correlation. (a) FDPE; (b) GB; (c) DDD.

model. The greatest differences are seen at the tips of the loops, presumably because of the large-scale motions of the loops (described earlier). Interestingly, the carbonyl carbon and oxygen atoms are particularly mobile in the DDD simulations, resulting in small spikes in the fluctuation plot in Figure 3.

The RMS deviations of each simulation from the starting structure are plotted in Figure 4. The DDD model yields a nearly continuous increase in deviation throughout the simulation, with a maximum of about 3.3 Å; large jumps in the RMSD can be identified with the loop transitions visible in Figure 2. The deviations of the other two models are more modest, although it is true that an even longer simulation might show further change. Note, however, that a small RMSD is not an indicator of success in this study, because the active site flaps are known to shift when the inhibitor is removed.

Analysis of correlated motions. Another view of atomic motions is provided by cross-correlations between the positions of backbone atoms [eq. (6)]. These correlations are plotted in Figure 5a–c for the FDPE, GB, and DD simulations, respectively. Only correlations with absolute values ≥ 0.2 are shown in order to emphasize the important correlations and anticorrelations. A darker square indicates a stronger correlation (lower right triangle) or anticorrelation (upper left triangle). The most striking result is the similarity of the FDPE and GB correlation graphs to each other, and the dissimilarity of the DDD graph. In particular, the FDPE and GB models show a strong correlation between the tips of the two flaps that is absent in the DDD model. The weak correlation with the DDD model presumably results from the loss of hydrogen bonding between the flaps. There is almost no anticorrelation greater than 0.2 for the FDPE and GB simulations, but the



(b)

FIGURE 5. (Continued)

DDD simulation, in contrast, yields a substantial anticorrelation between the tip and the β-sheet body of flap 1.

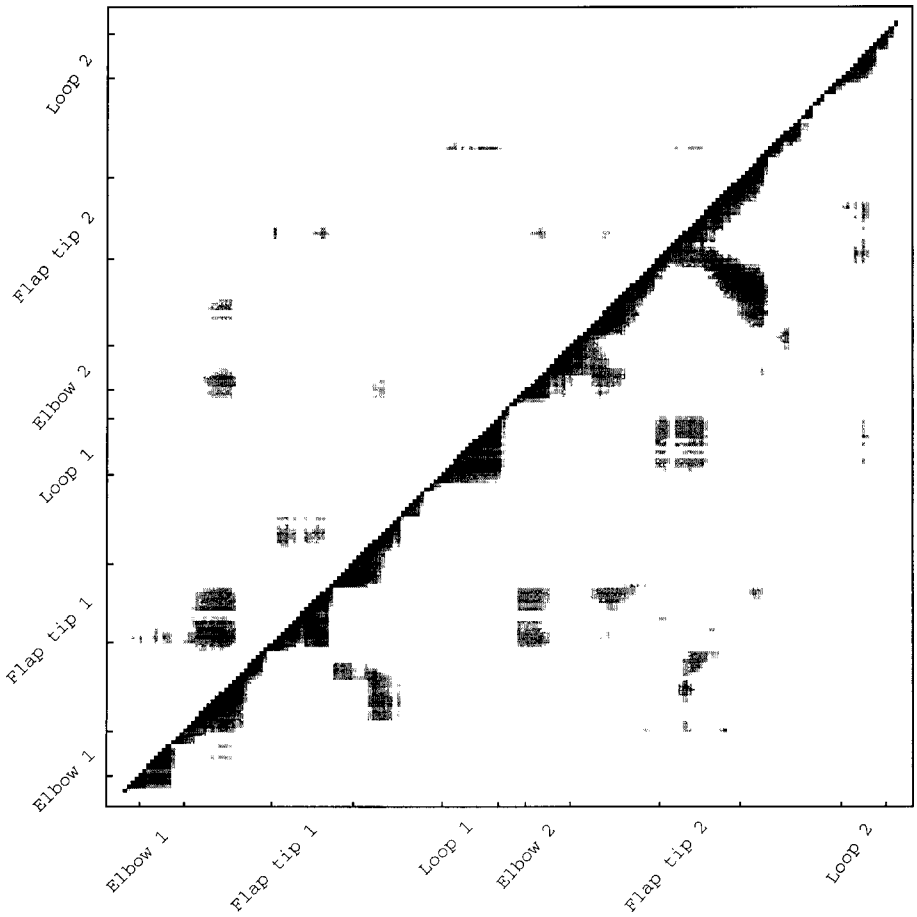
Temperature and energy. Temperature and energy data are summarized in Table VII. The temperature is fairly stable and near the set value of 300 K for all three simulations. The mean temperatures lie within the RMS fluctuation of this variable, but both the temperature—effectively the kinetic energy—and the total energy show weak negative drifts.

Discussion and Conclusions

The studies described here characterize the performance of various approximations to a detailed treatment of continuum electrostatics in calculations of energy, force, and motion. Of particular interest are the three matched 1-ns simulations of the active site of HIV-1 protease that treat electrostatic solvation via finite-difference solutions of the Poisson equation, the generalized Born approximation, and the distance-dependent dielectric model.

The GB model yields far more accurate energies and forces than the distance-dependent dielectric model, and its accuracy can be increased by using a few detailed finite-difference PE solutions to reparameterize it for the specific system under study. Unexpectedly, the GB model yields more accurate energies than forces, as judged by correlation coefficients and linear regression slopes. The accuracy of the GB forces is not greatly reduced by the time-saving approximation of neglecting the differentials of the effective Born radii with respect to atomic positions. This approximation is therefore used in the dynamics simulation with GB.

The distance-dependent dielectric model is markedly less accurate than the GB model. Indeed, forces calculated with this model for HIV-1 protease show a correlation coefficient of 0.4 relative to detailed finite-difference calculations. It should be possible to improve its accuracy by reparameteriz-



(c)

FIGURE 5. (Continued)

ing for the systems studied, as done for GB, but this is not pursued here.

Finite-difference solutions of the Poisson equation with a coarse grid (1 Å) yield energies that are somewhat less accurate than GB for a small molecule, but more accurate for the active site of HIV-1 protease. Forces computed with a 1-Å grid are considerably more accurate than those computed with GB, both for isoleucine and for HIV-1 protease, and a 1-Å grid is used in the dynamics simulation with FDPE.

The most significant result from the simulations of HIV-1 protease with an empty active site is that the distance-dependent dielectric model leads to collapse of the mobile loops into the active site, whereas the loops remained extended in both the GB and FDPE simulations. The collapse seen in the DDD simulation seems to be unrealistic, given that

TABLE VII.
Temperature and Energy Summaries for the Three Simulations.

Model	T (K)	T Fluct. (K)	T Slope (K/ps)	E^{tot} Slope (kJ/mol·ps)
FDPE	302.2	10.1	$-7 \cdot 10^{-4}$	−0.267
GB	294.8	9.8	$-4 \cdot 10^{-4}$	−0.130
DDD	295.7	9.9	$-7 \cdot 10^{-4}$	−0.125

Temp., temperature; T Fluct., root-mean-squared variation around the mean temperature; T Slope, first derivative of linear regression fit of temperature vs. time; E^{tot} Slope, first derivative of linear regression fit of the total energy vs. time.

the loops are believed to open up in the absence of ligand.[49] The unique behavior of the loops in the DDD simulation is consistent with the very low correlation between forces computed with this model and with GB and FDPE. Although the loops did not open up in the GB or FDPE simulations, the conformations sampled seem at least reasonable and longer simulations might finally show opening.

In summary, although using the DDD model is probably preferable to complete neglect of the electrostatic influence of water, it is a very crude approximation whose chief virtue is computational speed. The GB approximation is appealing because it is faster than FDPE and captures much of the same physics. It is not yet clear whether GB is less realistic than FDPE, but we do find that correcting GB toward FDPE does tend to improve agreement with experiment.[37] Thus, it seems preferable to use FDPE when it is not too slow for the application at hand. Advances in numerical methods for solving the FDPE[55–57] may make make this approach increasingly attractive. It is worth noting that the accuracy of GB can be improved by using a few detailed FDPE calculations to reparameterize it for the system under study, as demonstrated here.[37] Finally, systematic efforts to improve the accuracy of GB models are clearly of interest, especially for applications to macromolecules.

Acknowledgment

Certain commercial equipment or materials are identified in this article in order to specify the methods adequately. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. Kang, Y. K.; Gibson, K. D.; Némethy, G.; Scheraga, H. A. J Phys Chem 1988, 92, 4739.
2. Eisenberg, D.; McLachlan, A. D. Nature 1986, 319, 199.
3. Clark Still, W.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
4. Cramer, C. J.; Ruhlar, D. G. J Comput Chem 1992, 13, 1089.
5. Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. Molec Simul 1993, 10, 97.
6. Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Biochemistry 1994, 33, 13977.
7. Abagyan, R.; Totrov, M. J Mol Biol 1994, 235, 983.
8. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
9. Schmidt, A. B.; Fine, R. M. Molec Simul 1994, 13, 347.
10. Smith, P. E.; Pettitt, B. M. J Phys Chem 1994, 98, 9700.
11. Roux, B.; Simonson, T. Biophys Chem 1999, 78, 1.
12. Lazaridis, T.; Karplus, M. Prot Struct Funct Genet 1999, 35, 133.
13. Gilson, M. K.; Rashin, A. A.; Fine, R.; Honig, B. J Mol Biol 1985, 183, 503.
14. Warwicker, J.; Watson, H. C. J Mol Biol 1982, 157, 671.
15. Gilson, M. K.; Honig, B. Prot Struct Funct Gen 1988, 4, 7.
16. Zauhar, R. J.; Morgan, R. S. J Mol Biol 1985, 186, 815.
17. Yoon, B. J.; Lenhoff, A. M. J Comput Chem 1990, 11, 1080.
18. Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Comput Phys Commun 1991, 62, 187.
19. You, T. J.; Harvey, S. C. J Comput Chem 1993, 14, 484.
20. Zauhar, R. J.; Morgan, R. S. J Comput Chem 1990, 11, 603.
21. Davis, M. E.; McCammon, J. A. J Comput Chem 1991, 12, 909.
22. Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. J Phys Chem 1993, 97, 3591.
23. Gilson, M. K.; McCammon, J. A.; Madura, J. D. J Comput Chem 1995, 16, 1081.
24. Smart, J. L.; Marrone, T. J.; McCammon, J. A. J Comput Chem 1997, 18, 1750.
25. Wan, S. Z.; Wang, C. X.; Xiang, Z. X.; Shi, Y. Y. J Comput Chem 1997, 18, 1440.
26. Im, W.; Beglov, D.; Roux, B. Comput Phys Commun 1998, 111, 59.
27. Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. Recl Trav Chim Pays-Bas 1956, 75, 487.
28. Peradejordi, F. Cah Phys 1963, 17, 393.
29. Jano, I. Comput Rend Acad Sci (Paris) 1965, 261, 103.
30. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Chem Phys Lett 1995, 246, 122.
31. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Clark Still, W. J Phys Chem A 1997, 16, 3005.
32. Schaefer, M.; Bartels, C.; Karplus, M. J Mol Biol 1998, 284, 835.
33. Dominy, B. N.; Brooks, C. L., III J Phys Chem B 1999, 103, 3765.
34. Jayaram, B.; Liu, Y.; Beveridge, D. L. J Chem Phys 1998, 109, 1465.
35. Ghosh, A.; Sendrovic Rapp, C.; Friesner, R. A. J Phys Chem B 1998, 102, 10983.
36. Gilson, M. K.; Honig, B. J Comput Aided Molec Design 1990, 5, 5.
37. Luo, R.; Head, M. S.; Moult, J.; Gilson, M. K. J Am Chem Soc 1998, 120, 6138.
38. Luo, R.; David, L.; Hung, H.; Devaney, J.; Gilson, M. K. J Phys Chem B 1999, 103, 727.
39. Luo, R.; Head, M. S.; Given, J. A.; Gilson, M. K. Biophys Chem 1999, 78, 183.
40. David, L.; Luo, R.; Head, M. S.; Gilson, M. K. J Phys Chem B 1999, 103, 1031.
41. Given, J. A.; Gilson, M. K. Proteins 1998, 33, 475.
42. Kiso, Y. Biopolymers (Peptide Sci) 1996, 40, 235.
43. QUANTA, Molecular Simulations, Inc., Waltham, MA.

44. Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure 1995, 3, 581.
45. Wang, Y.-X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Torchia, D. A. Biochemistry 1996, 35, 9945.
46. Stratton, J. A. Electromagnetic Theory; McGraw-Hill: New York, 1941.
47. Hyland, L. J.; Tomaszek, T. A., Jr.; Meek, T. D. Biochemistry 1991, 266, 24349.
48. van Gunsteren, W. F.; Berendsen, H. J. C. Mol Simul 1988, 1, 173.
49. Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J. Biochimie 1991, 73, 1391.
50. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminanathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
51. Luty, B. A.; Davis, M. E.; McCammon, J. A. J Comput Chem 1992, 13, 768.
52. Harte, W. E.; Swaminathan, S.; Mansuri, M. M.; Martin, J. C.; Rosenberg, I. E.; Beveridge, D. L. Proc Natl Acad Sci USA 1990, 87, 8864.
53. Venable, R. M.; Brooks, B. R.; Carson, F. W. Prot Struct Funct Genet 1993, 15, 374.
54. Richards, F. M. Annu Rev Biophys Bioeng 1977, 6, 151.
55. Nicholls, A.; Honig, B. J Comput Chem 1991, 12, 435.
56. www.eyesopen.com.
57. Luo, R.; David, L.; Gilson, M. K., in preparation.